



# A Lexicon of French Quotation Verbs for Automatic Quotation Extraction

Benoît Sagot, Laurence Danlos, Rosa Stern

## ► To cite this version:

Benoît Sagot, Laurence Danlos, Rosa Stern. A Lexicon of French Quotation Verbs for Automatic Quotation Extraction. 7th international conference on Language Resources and Evaluation - LREC 2010, May 2010, Valetta, Malta. inria-00515461

**HAL Id: inria-00515461**

**<https://inria.hal.science/inria-00515461>**

Submitted on 7 Sep 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Lexicon of French Quotation Verbs for Automatic Quotation Extraction

Benoît Sagot<sup>1</sup>, Laurence Danlos<sup>1</sup>, Rosa Stern<sup>1,2</sup>

1. Alpage, INRIA Paris-Rocquencourt & Université Paris 7, 30 rue du Château des Rentiers, 75013 Paris, France

2. Agence France-Presse – Medialab, 2 place de la Bourse, 75002 Paris, France

benoit.sagot@inria.fr, laurence.danlos@linguist.jussieu.fr, rosa.stern@afp.com

## Abstract

Quotation extraction is an important information extraction task, especially when dealing with news wires. Quotations can be found in various configurations. In this paper, we focus on direct quotations introduced by a parenthetical clause, headed by a “quotation verb”. Our study is based on a large French news wire corpus from the Agence France-Presse. We introduce and motivate an analysis at the discursive level of such quotations, which differs from the syntactic analyses generally proposed. We show how we enriched the *Lefff* syntactic lexicon so that it provides an account for quotation verbs heading a quotation parenthetical, especially those extracted from a news wire corpus. We also sketch how these lexical entries can be extended to the discursive level in order to model quotations introduced in a parenthetical clause in a complete way.

## 1. Introduction

Information sourcing constitutes a major part of press agencies’ work. In particular, explicit quotations are massively represented in news wires. In this context, information extraction techniques for retrieving quotations are particularly useful in order to synthesize information on a given topic, including past related quotations from relevant people. The main requirement for such techniques is an exhaustive coverage of various forms of quotations, their structures and their lexical properties. Our study focuses on French quotations as found in the corpus of AFP (Agence France Presse, the main French press agency).

(French) Quotations can be found in the following configurations:

**IQ** *Indirect quotation* embedded within the sentential complement of a “reported speech verb”; the example in (1a) illustrates the case of “mixed quotation” (Davidson, 1979);

**DI** *Direct quotation* with the verb that introduces it in *initial* position, (1b);

**DP** *Direct quotation* with the verb that introduces it in a *parenthetical* clause in median or final position, (1c); it must be underlined that the subject inversion is mandatory in French; the parenthetical is marked as a “quotation parenthetical”.

- (1) a. Le sénateur a déclaré (à la ministre) qu’il allait “défendre cette loi contre vents et marées.”  
*The senator declared (to the minister) that he would “defend this law despite all the obstacles.”*
- b. Le sénateur a déclaré (à la ministre) : “Je vais défendre cette loi contre vents et marées.”  
*The senator declared (to the minister): “I will defend this law despite all the obstacles.”*
- c. “Je vais défendre cette loi contre vents et marées”, a déclaré le sénateur (à la ministre).  
*“I will defend this law despite all the obstacles”, the senator declared (to the minister).*

Quotations have been extensively studied in the literature. Since we aim at building an automatic quotation extraction tool, we concentrate here only on the following questions about quotations: which are the (French) verbs that introduce a quotation in an IQ, DI and/or DP configuration (such verbs are called “quotation verbs”)? Which lexical entry should be given to a quotation verb in a syntactic lexicon used by parsers, possibly within a quotation extraction system?

To answer these questions, we semi-automatically built an inventory of quotation verbs by exploring the AFP corpus (section 2.). This inventory lead us to propose an analysis of the relation between a quotation and the quotation verb within a DP that is different from the main analyses proposed in French literature (section 3.). Indeed, we consider the semantic relation between the quotation and the quotation parenthetical as located at the discursive level: it can not be obtained by a simple sentence-level syntax-semantic interface. We integrated the lexical syntactic consequences of this analysis within the *Lefff*, a syntactic lexicon for French which is used at different stages in the SAPIENS quotation extraction system (Villemonde de La Clergerie et al., 2009) (see Section 4.). Indeed, SAPIENS relies on the results of a deep parsing system based on the French FRMG parser (Thomasset and Villemonde de la Clergerie, 2005) which relies on the *Lefff*. Finally, we model the discursive consequences of our analysis of DP with discursive subcategorization frames (see Section 5.).

## 2. Corpus-based study of quotations

There exists a fair amount of (French) prototypical reported speech verbs, such as speech verbs (*dire/say, déclarer/declare*) and verbs of propositional attitude (*craindre/fear, regretter/regret*). Lists of such verbs are available in the literature, that are based on corpus studies and/or on semantic criteria (Lamiroy and Charolles, 2008; Monville-Burston, 1993). We already had such a list of 110 verbs at our disposal, drawn up before the study described here, mostly as a side-effect of the development of the FRMG parser.

The first step of our corpus-based study of quotations was aimed at extracting quantitative information from our news wire corpus, in order to evaluate the relative importance of the three quotation configurations listed above.

We first looked in a corpus of 5000 AFP news wires for quotation configurations (IQ, DI, DP) involving inflected forms of quotation verbs from our initial list. We achieved this by means of a set of regular expressions meant to get a broad coverage and a good detection recall. In consequence, out of the 2,742 candidates we gathered, many of them did not correspond to quotations. Therefore, we sorted out the results manually, thus building a corpus of 836 quotation configuration occurrences associated with quotation verbs. Out of the 110 verbs in our initial list, 91 are represented.

A manual classification of these occurrences among the three quotation configurations allowed us to make the following observations:

- a vast majority of the verbs occur in DPs: 70 over 91;
- among these 70 verbs, 20 occur only in DPs; the latter are either pronominal verbs, intransitive verbs or verbs not usually listed as reported speech or propositional attitude verbs;
- the distribution among all verbs between IQs, DIs and DPs shows that DPs are the most frequent configurations, with respectively 20.6%, 2.9% and 65.8%.<sup>1</sup>

These three observations suggest that DP is a somewhat productive configuration. We therefore decided to leave aside the case of DIs for further work and focus on DPs. This led us to try and gather a more comprehensive list of quotation verbs that can occur in DPs and their associated properties.

In order to carry out this second step of our corpus-based study, we applied on the corpus the subset of the regular expressions developed during the first step that represent solely DPs. The result is an extended list of 232 verbs associated with DP occurrences. We then included all newly extracted verbs in the complete set of regular expressions (including those for IQs and DIs). This allowed us to manually classify the extended verb list, once again w.r.t. their distribution over IQs, DIs and DPs. This showed the following:

- any verb occurring in an IQ can also occur in a DP;
- more verbs extracted during the second phase only occur in DPs; moreover, as already found out during the first step of our corpus-based study, such verbs are either pronominal, intransitive or transitive but not prototypical reported speech verbs; overall, only one half of the 232 verbs are reported speech verbs which can introduce an IQ.

<sup>1</sup>The remaining 10.8% of the occurrences correspond to “pure quotations”, where the verb does not grammatically support the quotation, as in *Un récent article jugeait cette candidature “scandaleuse”, et accusait M. Hosni. (A recent article considered this candidacy as “disgraceful”, and accused Mr. Hosni.)* We do not deal with this case here since we are interested in quotation verbs.

After a more careful analysis of these 232 verbs, we have divided them up into three classes:

- Class 1: 118 transitive reported speech verbs such as *dire/say* or *répondre/answer* which appear in both IQs and DPs,
- Class 2: 43 intransitive verbs such as *ricaner/snigger* or *fulminer/fulminate* which appear in DPs but not in IQs,
- Class 3: 71 transitive verbs such as *commenter/comment* or *continuer/continue* which appear in DPs but not in IQs.

These surprising results, especially concerning Class 3, require an analysis on DPs and verbs that head their parenthetical clauses, conducted in the following section.

### 3. Analysis of quotation verbs in DPs

In French studies, it is commonly stated that the quotation in a DP is linked to the *direct object* of the quotation verb. For example, (Bonami and Godard, 2008) consider that the quotation has the characteristics of an *extracted object*<sup>2</sup>. However, their study is based only on reported speech verbs.

(Lamiroy and Charolles, 2008) consider both (transitive) reported speech verbs of Class 1 and intransitive verbs of Class 2. They put forward the hypothesis that “speech verbs” are basically transitive (with the speech argument realized as the direct object). Therefore they define a transitivity scale: *ricaner/snigger* is “less” transitive than *dire/say* since it can be a quotation verb in a DP, (2a), but not in an IQ, (2b).

- (2) a. “Je m’opposerai à cette loi”, ricana le sénateur.  
*“I will oppose this law”, the senator sniggered.*
- b. \*Le sénateur a ricané qu’il s’opposerait à cette loi.  
*The senator sniggered that he would oppose this law.*

It doesn’t seem justified to consider *ricaner/snigger* as transitive just because it can be the head of a quotation parenthetical. Moreover, the hypothesis advocated in (Lamiroy and Charolles, 2008) doesn’t hold for transitive verbs of Class 3, whose direct object doesn’t refer to the speech argument, as we are going to show.

When no quotation is involved, the verb *commenter* takes an obligatory direct object and an optional oblique object introduced by *par/with*, (3a). The direct object refers to the commented thing — this semantic argument is noted *X* —, while the oblique object refers to the commentary — noted *Y*.<sup>3</sup> These two complements are compulsory nominal: (3b) and (3c) are unacceptable. Therefore, *commenter* cannot head an IQ configuration. On the other hand, *commenter* can be a quotation verb in a DP, (3d) — and it has the same

<sup>2</sup>They model the extracted object in HPSG with a *slash* feature in the parenthetical clause.

<sup>3</sup>When *Y* is not syntactically realized, it is qualified as “unspecified”.

meaning in (3a) and (3d). In (3d), the commentary  $Y$  is given in the quotation, while  $X$  is given in the left discursive context: it corresponds to the government's decision of cutting taxes (which is the antecedent of *ce/it*).

- (3) a. Le député a commenté cette décision (par une remarque acerbe).  
*The deputy commented on this decision (with a caustic remark).*
- b. \*Le député a commenté que les impôts allaient être réduits (par une remarque acerbe).  
*The deputy commented (on) that taxes will be cut (with a caustic remark).*
- c. \*Le député a commenté cette décision par (ce) qu'il trouvait cela anticipé.  
*The deputy commented on this decision with that he found it premature.*
- d. Le gouvernement a annoncé sa décision de réduire les impôts. "C'est prématuré", a commenté le député.  
*The government announced their decision to cut taxes. "It is premature", the deputy commented.*

The quotation  $Y$  in (3d) is not linked to the direct object of *commenter* since the latter refers to  $X$ . Moreover, the semantic argument  $X$  cannot be realized in the parenthetical (neither as an NP: \**a commenté le député cette décision* nor as a clitic: \**l'a commentée le député*) although it cannot be left unspecified: it must be retrieved in the left discursive context of the DP. This left context cannot be empty: a discourse cannot start with a DP whose quotation verb is *commenter*, even if the quotation doesn't include any anaphoric element as in (4), which is incoherent (hence the sign #) in an empty left context.

- (4) # "Les paysans vont descendre dans la rue", commenta le député.  
*"Farmers are going to take to the streets", the deputy commented.*

In a nutshell, *commenter* has three semantic arguments:  $x$  the commentator,  $X$  what is commented, and  $Y$  the commentary. When no quotation is involved,  $x$  is realized as the subject,  $X$  as a direct object and  $Y$  as an optional oblique object. These data can be recorded in a standard subcategorization frame which states both syntactic and semantic constraints and the mapping between syntactic and semantic arguments. When *commenter* is used as the head of a parenthetical in a DP, only  $x$  is realized in the parenthetical;  $X$  must be retrieved in the left discursive context of the DP and  $Y$  corresponds to the quotation with no syntactic link with *commenter*. These data cannot be recorded in a standard subcategorization frame which states only constraints falling within the sentential level. They require to be recorded in two subcategorization frames:

- one at the sentential level which states the syntactic constraints,

- one at the discursive level which states the semantic constraints.

As the link between the quotation and the parenthetical in a DP such as (3d) is not syntactic, this DP — as well as other DPs whose quotation verb is a transitive verb of Class 3 — should be considered as a "discursive construction" in which syntax plays a role only to state that the argument  $x$  of *commenter* must be realized as the subject — which must be inversed in French —, and to state constraints on the position of the parenthetical within the DP. Similarly, DPs whose quotation verb is an intransitive verb of Class 2 (*ricaner/snigger*) should be considered as constructions with no syntactic link between the quotation and the parenthetical. For DPs whose quotation verb is a transitive verb of Class 1, two solutions can be contemplated:

- either the quotation is syntactically linked to the parenthetical: it is an extracted object as advocated in (Bonami and Godard, 2008);
- or the quotation and the parenthetical forms a discursive construction, without any syntactic link.

We adopt the second solution for three reasons. First, it allows an homogeneous analysis of all DPs, no matter which class the quotation verb belongs to. Second, it allows us to take into account the fact that a DP can be embedded neither under a verb with a clausal argument, (5a), nor under a subordinating conjunction, (5b).

- (5) a. \*Je crois que "donne-moi la main !" a lancé Paul à Marie. [(Bonami and Godard, 2008)]  
*I think that "give me your hand!", Paul called out to Mary.*
- b. \*Marie est furieuse parce que "Tu es paresseuse", lui a dit Fred.  
*Mary is angry because "You are lazy", Fred said to her.*

On the other hand, a sentence with an extracted object in a cleft clause can be embedded, (6). Therefore, the phenomenon of extraction doesn't explain the embedding interdiction.

- (6) a. Je crois que c'est Fred que Sue aime.  
*I think that it is Fred that Sue loves.*
- b. Marie est furieuse parce que c'est Fred que Sue aime.  
*Mary is angry because it is Fred that Sue loves.*

This interdiction for DPs to be embedded, which is also observed for multi-sentential discourses, strongly indicates that DPs fall within the discursive level and not the syntactic-sentential level.

Third, the second solution allows us to take into account the following data: some reported speech verbs of Class 1 such as *répondre/answer* or *ajouter/add* share with verbs

of Class 3 the impossibility of heading a quotation parenthetical in a DP which occurs in an empty left discursive context: (9a) is incoherent contrarily to (9b).<sup>4</sup>

- (9) a. #“Les paysans vont descendre dans la rue”,  
répondit le député.  
“Farmers are going to take to the streets”, the  
deputy answered.
- b. Le Premier ministre a demandé comment les  
paysans allaient réagir à cette loi. “Les paysans  
vont descendre dans la rue”, répondit le député.  
The Prime minister asked how farmers are going  
to respond to this law. “Farmers are going to take  
to the streets”, the deputy answered.

In conclusion, we propose the hypothesis that all DPs — whatever the class of the quotation verb involved — are discursive constructions, which require two subcategorization frames for the quotation verb.

Postulating two subcategorization frames is somehow unusual since it is generally considered that a lexical item has a single subcategorization frame which states both syntactic and semantic constraints, these constraints falling within the sentential level. However, adverbial discourse connectives such as *ensuite/then* or *de ce fait/therefore* also require two subcategorization frames: at the syntactic level, they have a single argument<sup>5</sup>, while at the semantic level, they

<sup>4</sup>However, *répondre/answer* and *commenter/comment* differ on the following point: the argument *X* of *répondre* (what is answered) can be syntactically realized in the parenthetical as an anaphoric NP, (7), while it should be reminded that the argument *X* of *commenter* (the thing commented) cannot be realized in the parenthetical at all.

- (7) a. Le Premier Ministre a demandé comment les paysans  
allaient réagir à cette loi. “Les paysans vont descendre  
dans la rue”, répondit le député à cette question  
saugrenue.  
The Prime Minister asked how farmers are going to re-  
spond to this law. “Farmers are going to take to the  
streets”, the deputy answered this ridiculous question.
- b. Le Premier Ministre a demandé comment les paysans  
allaient réagir à cette loi. “Les paysans vont descendre  
dans la rue”, lui répondit le député.  
The Prime Minister asked how farmers are going to re-  
spond to this law. “Farmers are going to take to the  
streets”, the deputy answered to him.

Therefore, it should be considered that the argument *X* of *répondre* is realized as an anaphor  $\emptyset$  in (9b). The realization of an argument as an anaphor  $\emptyset$  is observed in other contexts, for example in (8), (Danlos, 2005)

- (8) Cet été, Fred a entrepris de repeindre toute sa maison. Il a  
commencé  $\emptyset$ /cette entreprise par le salon.  
This summer, Fred undertook to repaint his whole house. He  
started  $\emptyset$ /this enterprise with the living-room.

On the other hand, the fact that the argument *X* of *commenter* can never be realized in a quotation parenthetical is modelled by withdrawing the corresponding syntactic function, see Section 4..

<sup>5</sup>This argument is of category S if the connective appears in initial position (*Ensuite, Fred a décidé de partir/Then, Fred has decided to leave*) and of category V if the connective appears in-

have two arguments, one of which must be retrieved in the left discursive context of the sentence the connective appears in. This discrepancy between the number of syntactic and semantic arguments can be taken into account with two subcategorization frames, as we propose for quotation verbs in DPs.

Postulating two subcategorization frames, one at the syntactic-sentential level, the other one at the semantic-discursive level, requires designing a non trivial syntax-semantics interface based on a sentence-discourse interface. It is not in the scope of this paper to discuss sentence-discourse interfaces<sup>6</sup>. However, we introduce in the next section the lexical entries for quotation verbs in DPs with their sentential subcategorization frame(s), and in the following one their discursive counterparts.

#### 4. Lexical entries for quotation verbs with their sentential subcategorization frames

We integrated the lexical data collected in our corpus study (section 2.) with the analysis sketched in the previous section in a lexicon, the *Lefff* (Lexique des formes fléchies du français — *Lexicon of French inflected forms*) (Sagot, 2010). The *Lefff* is a freely available large-coverage lexicon that relies on the Alexina lexical formalism. Along morphological information, it associates syntactic information with each “intensional” entry of a given lemma, each of them corresponding to a different meaning of this lemma. Syntactic information includes, among other, a deep subcategorization frame (syntactic functions and their possible realizations at the deep syntactic level) and acceptable redistributions that generate automatically various surface subcategorization frames (for verbs, possible redistributions include %active, %passive, %impersonal, etc.). The lexical model of the *Lefff*, Alexina, is formalism-independent, which allows Alexina lexicons to be used in parsing systems relying on various formalisms, in particular (lexicalized) TAGs (e.g., FRMG).

When a verb is used as the head of a quotation parenthetical, it has the same meaning as the meaning it has (or one of the meanings it can have) when it governs an IQ or when it is used in contexts without quotations. For example, let us remind that *commenter/comment* has the same meaning in (3a) and (3d). Moreover, as underlined by (Bonami and Godard, 2008), the quotation verb imposes selectional restrictions on the type of speech act that the quotation may express, and these restrictions are compatible with those observed in contexts without quotations. For instance, the verb *demande/ask*, when used as a quotation verb, requires the quotation to be an interrogative (10a), which corresponds to the fact that it can be used with an indirect interrogative (10b). On the contrary, *affirmer/affirm* is only built with assertions (10c,10d).

- (10) a. “Quelle heure est-il ?”, demanda-t-il.  
“What time is it?”, he asked.

side the VP (*Fred a ensuite décidé de partir/Fred has then decided to leave*).

<sup>6</sup>In a discourse understanding perspective, sentence-discourse interfaces are discussed in (Webber, 2004) in the framework of D-LTAG and in (Danlos, 2009) in the framework of D-STAG.

b. Il demanda quelle heure il était.  
He asked what time it was.

c. “Il est 2h30”, affirma-t-il.  
“It is 2:30”, he affirmed.

d. Il affirma qu’il était 2h30.  
He affirmed that it was 2:30.

For these reasons, we model the possibility for a verb to head a quotation parenthetical in a DP by adding to its existing entry a new redistribution. This redistribution, prefixed by %DP, is defined as follows:

- it removes all syntactic functions that cannot be realized in a parenthetical, namely
  - in all cases, the argument that corresponds to the quotation itself (this has no effect on Class 2 verbs);
  - for Class 3 verbs, the direct object (Obj), as explained in Section 3..

This leads to various %DP redistributions, depending on the class and on the syntactic function corresponding to the quotation; e.g., %DP\_noObj\_noObl, associated with some Class 3 verbs such as *commenter/comment* removes the Obl because it corresponds to the quotation as well as the Obj;

- it discards any sentential (scompl) or infinitive (sinf) realization for the subject (Suj),
- it adds a special feature [DP=+] that is meant to be used in the grammar to force the subject inversion and identify the presence of a quotation parenthetical;
- it adds a feature that represents the selectional restrictions affecting the quotation, such as [assertion=+] or [interrogative=+]; this allows the output of the parser to be compatible with a further discursive analysis.

For example, (simplified) *Lefff* entries for the verbs *dire/dit* (class 1), *ricaner/snigger* (class 2) and *commenter/comment* (class 3) are shown respectively in (11a), (11c) and (11e). Each of these “intensional” entries yields many different extensional entries, for each inflected forms and redistributions. For each of these verbs, we provide the (simplified) DP extensional entry for the third person singular *passé simple* in respectively (11b), (11d) and (11f). For example, the DP redistribution %DP\_noObj yields an entry for *dit/say* in which the Obj is removed but the optional Objà is preserved, since it can be realized in the parenthetical, see (1c).

- (11) a. *dire* Suj:cln|sn,  
Obj:cla|sn|de-sinf|scompl|qcompl,  
Objà:(cld|à-sn);;  
%active,%passive,%se\_moyen,  
%impersonnal\_passive,%DP\_noObj
- b. *dit* Suj:cln|sn,Objà:(cld|à-sn);DP=+,assertion=+;
- c. *ricaner* Suj:cln|sn;;%active,%DP\_intrans

d. *ricana* Suj:cln|sn;DP=+,assertion=+;

e. *commenter* Suj:cln|sn,Obj:cla|sn,  
Obl:(par-sn);;  
%active,%passive,%se\_moyen,  
%DP\_noObj\_noObl

f. *commenta* Suj:cln|sn;DP=+,assertion=+

## 5. Discursive subcategorization frames

In this section, we sketch the discursive/semantic subcategorization frames for quotation verbs when they head a quotation parenthetical in a DP (as usual for the discursive level, these frames include discourse relations linking Elementary Discourse Units (EDUs in short)).

For verbs of Class 3 such as *commenter/comment*, the semantic/discursive subcategorization frame can be given in the (simplified) formula:

$$(12) \lambda x Y X. \exists W (W \cong (\exists e. \text{speechAct}(e, x)) \wedge \text{Attribution}(W, Y) \wedge \text{Commentary}(Y, X)) \\ \text{with } x :: e, X, Y :: t$$

In words, there exists an entity  $x$ , the referent of the subject of *commenter*, who performed a speech act  $e^7$ . This event forms the core of an EDU,  $W$ .<sup>8</sup>  $W$  is linked via the discourse relation *Attribution* (Redeker and Egg, 2006) to the quotation  $Y$ , which is itself linked via the relation *Commentary* (Asher and Lascarides, 2003) to an EDU  $X$  (which must be retrieved in the left discursive context of the DP), see (Danlos et al., 2010) for more details. For a verb of Class 3 different from *commenter/comment*, the discourse relation *Commentary* should be replaced by another discourse relation, e.g. *Continuation* for *continuer/continue*. For verbs of Class 1 such as *répondre/answer* or *ajouter/add* which can head a parenthetical quotation only in DPs which occur in a non empty left discursive context (see examples (9) in section 3.), the semantic/discursive subcategorization frame could be identical to that of *commenter* given in (12) except that the discourse relation *Commentary* is replaced by an appropriate relation, e.g. *Answer* for *répondre/answer* or *Continuation* for *ajouter/add*.

For the other verbs of Class 1 such as *annoncer/announce* or *dire/say*, the discursive frame could be the (simplified) following formula, in which the link with the left discursive context, if any, is left unspecified.

$$(13) \lambda x Y. \exists W (W \cong (\exists e. \text{annoncer}(e, x)) \wedge \text{Attribution}(W, Y))$$

Finally, for intransitive verbs of Class 2 such as *ricaner/snigger*, the semantic/discursive frame could be the one shown in (14), in which the EDU  $W$  consists of two overlapping events — the overlapping is represented as

<sup>7</sup>No modality of this speech act is specified: neither the manner of speaking, nor the instrument of communication if any, neither any non neutral propositional attitude.

<sup>8</sup>If the parenthetical includes adjuncts (*commenta le député après le dîner/the deputy commented after dinner*), their semantic representations are inserted in  $W$ .

$\circ(e, e_1)$ ) — one ( $e$ ) being a speech act performed by  $x$ , the other one ( $e_1$ ) being the fact that  $x$  was sniggering when performing her speech act.<sup>9</sup>

$$(14) \lambda xY. \exists W(W \cong (\exists ee_1.\text{speechAct}(e, x) \wedge \text{ricaner}(e_1, x) \wedge \circ(e, e_1)) \wedge \text{Attribution}(W, Y))$$

## 6. Conclusion and future work

We advocated an analysis at the discursive level of French DPs, which differ from the syntactic analyses generally proposed, and enriched the *Lefff* syntactic lexicon so that it can handle quotation verbs, especially those extracted from a news wire corpus. We also sketched how lexical entries can be extended to the discursive level in order to model in a complete way the DP phenomenon.

The next steps of this work are threefold. First, we intend to integrate the extended *Lefff* within the SAPIENS quotation extraction tool (Villemonde de La Clergerie et al., 2009). SAPIENS is an interactive platform for extracting and visualizing quotations from news wires, associated with their author and context. Its originality is that it relies on a deep linguistic processing chain, including the FRMG parser (Villemonde de La Clergerie, 2005). This allows for extracting quotations with a wide coverage and an extended definition, including quotations which are only partially quotes-delimited verbatim transcripts. But this also means that the integration of the new DP redistributions and features requires important changes in the (meta-)grammar on which FRMG is based.

On the longer term, we aim at extending the scope of our work by two means. First, we plan to describe in a systematic way the discursive analysis of quotation configurations. Second, we intend to work on the integration of other types of quotation configurations, namely DI and non-verbal constructions (e.g., constructions introduced by prepositions such as *selon/according to*).

## Acknowledgments

This work was supported in part by the Scribo project funded by the French “Pôle de compétitivité” System@tic (“Fonds Unique Interministériel”) and by the EDyLex project funded by the French National Research Agency (grant number ANR-09-CORD-008).

## 7. References

- Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press, Cambridge.
- Olivier Bonami and Danièle Godard. 2008. Syntaxe des incises de citation. In *Actes du Premier Colloque Mondial de Linguistique Française*, pages 209–225, Paris, France.
- Laurence Danlos, Benoît Sagot, and Rosa Stern. 2010. Analyse discursive des incises de citations. In *Actes du Second Colloque Mondial de Linguistique Française*, New Orleans, USA. To appear.

- Laurence Danlos. 2005. Partition of an entity with aspectuo-temporal operators. In *Proceedings of the Third International Workshop on Generative Approaches to the Lexicon (GL'2005)*, Geneva, Switzerland.
- Laurence Danlos. 2009. D-STAG: a formalism for discourse analysis based on SDRT and using synchronous TAG. In *Proceedings of the 14th Conference on Formal Grammar (FG'09)*, pages 1–20, Bordeaux, France.
- Donald Davidson. 1979. Quotation. In *Inquiries Into Truth and Interpretation*, pages 79–82. Oxford University Press, Oxford.
- Maurice Gross. 1975. *Méthodes en syntaxe*. Hermann, Paris, France.
- Béatrice Lamiroy and Michel Charolles. 2008. Les verbes de parole et la question de l’(in)transitivité. *Discours*, 2.
- Monique Monville-Burston. 1993. Les verba dicendi dans la presse d’information. *Langue française*, 98(1).
- Gisela Redeker and Markus Egg. 2006. Says who? on the treatment of speech attributions in discourse structure. In *Proceedings of the Constraints in Discourse Workshop (CID'06)*, Maynoth, Ireland.
- Benoît Sagot. 2010. The *Lefff*, a freely available and large-coverage morphological and syntactic lexicon for french. In *Proceedings of the 7th Language Resource and Evaluation Conference*, Valetta, Malta.
- François Thomasset and Éric Villemonde de la Clergerie. 2005. Comment obtenir plus des méta-grammaires. In *Proceedings of TALN'05*, Dourdan, France, June. ATALA.
- Éric Villemonde de La Clergerie, Benoît Sagot, Rosa Stern, Pascal Denis, Gaëlle Recourcé, and Victor Mignot. 2009. Extracting and visualizing quotations from news wires. In *Proceedings of L&TC 2009*, Poznań, Poland.
- Éric Villemonde de La Clergerie. 2005. From metagrammars to factorized TAG/TIG parsers. In *Proceedings of IWPT'05*, pages 190–191, Vancouver, Canada.
- Bonnie Webber. 2004. D-LTAG: extending lexicalized TAG to discourse. *Cognitive Science*, 28(5):751–779.

<sup>9</sup>This EDU  $W$  formalizes the non formal “fusion” analysis proposed in (Gross, 1975): *ricaner* used as the head of a quotation parenthetical is equivalent to “dire en ricanant” (say while sniggering).